

Internal Markets for Supply Chain Capacity Allocation

David McAdams and Thomas W. Malone

*Sloan School of Management
Massachusetts Institute of Technology*

MIT Sloan School of Management Working Paper No. 4546-05

MIT Center for Coordination Science Working Paper No. 224

June 2005

Internal Markets for Supply Chain Capacity Allocation

David McAdams and Thomas Malone, MIT Sloan*

Abstract

This paper explores the possibility of solving supply chain capacity allocation problems using *internal markets* among employees of the same company. Unlike earlier forms of transfer pricing, IT now makes it easier for such markets to involve many employees, fine-grained transactions, and frequently varying prices. The paper develops a formal model of such markets, proves their optimality in a baseline condition, and then analyzes various potential market problems and solutions. Interestingly, these proposed solutions are not possible in a conventional market because they rely on the firm's ability to pay market participants based on factors other than just the profitability of their market transactions. For example, internal monopolies can be ameliorated by paying internal monopolists on the basis of corporate, not individual, profits. Incentives for collusion among peers can be reduced by paying participants based on their profits relative to peers. Profit-reducing competition among different sales channels can be reduced by imposing an "internal sales tax". And problems caused by fixed costs can be avoided by combining "conditional internal markets" with a "pivot mechanism".

*This version June 10, 2005. David McAdams email: mcadams@mit.edu. Thomas Malone email: malone@mit.edu. This work was supported, in part, by Intel Corporation, National Science Foundation (Grant No. IIS-0085725), and the MIT Center for Coordination Science. We would especially like to thank the following participants in the development of a scenario for Intel Corporation that provided a motivating example for the work in this paper: Jim Rice, Mary Murphy-Hoye, Jay Hopman, Adlar Kim, Ben Koo, Jim Hines, John Quimby, George Herman, and Paolo Goncalves.

1 Introduction

Every company has to somehow solve the “supply chain capacity allocation problem”. Most broadly, this problem is: How and when to allocate resources to making different products and selling them to different customers? More specifically, this problem includes the following sub-problems:

Production: How many of each type of product to make at what times and places?

Sales: How to price these products, allocate them to different locations and customers, and motivate salespeople to sell them?

Many of these sub-problems have been analyzed separately (e.g., product mix problems, factory scheduling problems, pricing problems), but they are all interdependent.

In most large companies today, these problems are “solved” by time-consuming, complex, and expensive hierarchical processes involving plant managers, factory schedulers, strategic planners, product managers, marketing analysts, sales managers, and others. Sometimes these people use formal optimization models in their work; often they do not. In general, there is a pervasive feeling in many companies that these processes are much more cumbersome, inflexible, and expensive than they should (and could) be.

In this paper, we explore a different approach to this broad problem: using market-like processes inside the boundary of a single firm (“internal markets”) to help allocate manufacturing capacity and determine the prices, delivery dates, and product mix of the products that firms sell to their customers. Continuing reductions in the costs of information technology now make it feasible to implement such processes in many situations where it would have previously been prohibitively complex or costly to do so. Our hypothesis is that such processes, if implemented well, could result in much more profitable and timely decisions at much lower cost.

Of course, working out all the details of how to do this in practice is a very nontrivial task. In this paper, we provide some foundations for this endeavor by formulating a precise model of one class of internal markets and analyzing when such markets maximize overall firm profits. Section 2 briefly summarizes previous work on internal markets. Section 3 then presents the benchmark internal markets model. Section 4 is the heart of the paper, exploring when and why internal markets lead to outcomes that maximize firm profits. Furthermore, in contexts in which this benchmark does not perform well, we incorporate and analyze new features that can restore optimality. Among the reasons that the benchmark internal market may not perform well are: inept participants (though see discussion in Section 3.4), internal market power (see Section 4.1 which incorporates “cooperative wages”), internal market collusion (see Section 4.2 which incorporates “peer-group relative wages”), externalities (see Section 4.3 which incorporates “taxes”), and fixed costs (see Section 4.4 which incorporates both “conditional internal markets” and “pivot wages”). While the benchmark internal market is akin to a standard competitive market, many of our remedies for market failures have no analogue in the conventional theory of external markets. The key difference is that, in external markets, market participants’ wages depend only on the profitability of market transactions. In an internal market, on the other hand, the firm can leverage its ability to tailor employee’s wage payoffs to more than just internal market profit. Section 5 summarizes the key results of the paper and discusses some extensions.

2 Background

For many years, some companies have used internal *transfer pricing* where one division pays another for goods or services [e.g., Eccles (1985), Ackoff (1994), Mendelson (1985)]. The prices

for these internal transfers are typically established by a few senior executives and changed only annually or quarterly.

More recently, as IT has made the necessary communication much cheaper, there have been some experiments with much more fine-grained and dynamic *internal markets* (see summaries in Malone (2004a), Malone (2004b), Kambil and van Heck (2002)). In these markets, larger numbers of employees typically have smaller and more frequent transactions with each other, and prices are expected to change frequently as supply and demand change.

For example, British Petroleum used such a system to help reduce its greenhouse gas emissions [Kambil and van Heck (2002), Morgheim (2000)]. Each of its business units was given a target reduction amount, but the business units could then buy and sell emission “permits” among themselves, using a specially designed electronic trading system. For instance, a business unit that found it easy to reduce emissions more than its target could sell the excess permits to divisions that found it hard to meet their targets. In this way, the business unit managers found for themselves the most economically efficient opportunities for emission reduction wherever they were in the company without centralized intervention from senior executives.

There have also been a number of experiments with internal “prediction markets” where people in a company buy and sell predictions about possible future events. For example, Hewlett Packard has used such a system to predict future sales of its printers [Plott (2000), Chen and Plott (1998), Chen, Fine, and Huberman (2001, 2003, 2004)]. Employees who believed, for instance, that sales in December would be in the range of 1000-2000 units could buy “shares” of this prediction. If the predictions were accurate, the shares would each be worth \$1; if not, they would be worth \$0. Surprisingly, these internal prediction markets often provided more accurate predictions than the marketing department’s own forecasts.

A motivating example. In this paper, we explore the possibilities for internal markets to allocate capacity in supply chains. Our work was motivated, in part, by a scenario we developed with Intel Corporation for how they could determine which products to make at which times in which factories. The basic idea is to have an “internal futures market” where plant managers sell the rights to have different products available at different times in the future. Intel sales people would then buy these rights with the expectation of reselling the products to external customers.

As new information about supply and demand becomes available, the prices in this market could be continually varying. For instance, if a salesperson learns that a customer is planning to postpone a big order from January to March, this salesperson could sell futures for January and buy them for March, thus lowering the prices for January and raising them for March. The markets might also include internal “speculators” who keep the market more efficient by trading whenever they think prices are out of line with what they should be. Plant managers would try to sell as many shares as possible of their most profitable products without exceeding their overall capacity, where the internal market allows them to learn which products are most profitable. At the time they need to start production for a given delivery date, the shares they have sold for that date would determine which products they would actually produce.

The primary purpose of this paper is to explore analytically the conditions under which internal markets like this would result in allocations that optimize overall firm profits.

3 Benchmark Model

A firm has a set of G goods which are produced at F factories and sold through S sales channels at times $T = 1, 2, \dots$. These sales channels could be individual salespeople, distributors, or other

channels, but for simplicity we shall call them “stores”. The firm’s problem is to decide how much of each good g to produce at each factory f and to sell at each store s at each time T . Decisions at each factory and store are made by employees we shall call “factory managers” and “store managers”, respectively. These employees will be paid wages that depend, at least in part, on internal market outcomes.

For the rest of this section, we lay out our baseline model and discuss its key assumptions. In later sections of the paper, we shall relax some of these assumptions.

3.1 Costs and Revenues

Information: Each store and factory (“participant”) learns at some time prior to T what its own revenues or costs will be at time T . (We do not explicitly model the learning process.)

(A1) There is no fixed cost of production (or fixed revenue from sales).

We shall relax assumption **(A1)** in Section 4.4, when we allow for fixed costs of production.

(A2) There are no economies or diseconomies of scope in production or in sales, i.e. each factory’s total cost and each store’s total revenue is additively separable across goods.

(A2) combines two assumptions, that there are no scope effects on fixed costs or on marginal costs. For instance, there would be economies of scope in fixed costs if setting up for production of several products at once is less expensive than doing so separately, whereas there would be economies of scope in marginal costs if less labor was required to produce an additional unit of several products simultaneously than separately. While it simplifies the analysis, the assumption that there are no scope effects on *marginal costs* is not essential. In fact, when there are economies of scope in marginal costs, internal market trading has the extra benefit

of inducing employees to become less specialized and thereby increasing firm profits. Similarly, when there are diseconomies of scope in marginal costs, internal market trading increases firm profits by inducing employees to become more specialized. The assumption that there are no scope effects on *fixed costs* is more important. When we allow for fixed costs in Section 4.4, therefore, we will also allow for economies or diseconomies of scope in fixed costs.

(A3) There are no externalities in production or in sales. That is to say, each factory’s total cost depends only on what it produces and each store’s revenue depends only on what it sells.

We shall relax assumption **(A3)** in Section 4.3, when we allow each store’s revenue to depend on the total quantity sold by other stores as well as its own sales.

(A4) There is no storage and no delay between production and sales: quantity produced at time T can only be sold at time T .

Delivery delay and/or delivery cost would not change the analysis in any essential way, because buyers can take these additional factors into account when deciding what to buy, etc. On the other hand, storage can be important if it allows the firm to increase profits by smoothing out production. McAdams (2005) shows how an internal market can be adapted to exploit storage opportunities.

3.2 Internal Market Rules

Each factory manager or store manager trades on behalf of his factory or store. While production and sales only occur at integer times $T = 1, 2, \dots$, trading can occur at any time $t \in [0, \infty)$ with the proviso that the “market” for goods to be produced at time T closes at time T .

Internal market money accounts. Each participant has an internal market money account balance which starts at zero and may be positive or negative during the course of trading.

Orders and order execution. For each good g , time t , and future delivery time $T \geq t$, each participant may place, leave, or remove orders to buy and/or sell any quantity of shares of (g, T) for any quantity at any price. These orders may be either (a) limit orders which will not be executed beyond a certain price, or (b) market orders which will be executed at whatever the current market price is. At time t , there is a queue of orders to buy and to sell. Trades are executed from this queue in the natural way, with sellers who offer the lowest price being matched with buyers who offer the highest price. These trades are executed at the “market-clearing price” given the current offer queues, and employees’ internal market money accounts are debited or credited accordingly. See the online supplement for details (available at www.mit.edu/~mcadams/papers/im/onlineappendix.pdf).

3.3 Internal market profit and wages

The manner in which wages are set is part of the internal market design. As we shall see, wages need not depend directly on internal market profit, but it is important that the firm be able to measure internal market profit.

(A5) Each employee cares only about his or her wage and strictly prefers a higher wage. (Employees may have any risk preference.)

As a first approximation, the assumption that employees care only about their wage seems a reasonable one for our purposes. While employees certainly have other sorts of concerns (such as promotion opportunities, doing a good job for its own sake, etc..) most of these will reinforce the tendency toward good outcomes for the firm.

(A6) Some time before employees' wages are set, the firm observes each factory's *total* cost of production at time T and each store's *total* revenue from sales at time T .

Assumption (A6) is relatively weak since total economic cost and revenue only need to be observed *ex post* and marginal cost and marginal revenue need never be observed. Even so, there are some situations in which this assumption may be too strong. Recall that economic costs include the "opportunity costs" associated with alternative uses of production inputs but not the "sunk costs" that can not be avoided. A factory manager has the incentive to portray its economic costs as being as low as possible, in particular to conceal opportunity costs and to pass off true economic costs as sunk. For instance, the wages paid to factory workers that can not be fired or employed in any alternative use are a sunk cost and hence should not be included in total economic costs when it comes to determining that factory manager's wage. On the other hand, if those workers can be put to work producing some good that is not traded in the internal market, their wages should be counted against the factory manager as an opportunity cost. This gives factory managers an incentive not only to hide their alternative opportunities but to bias production in favor of goods that are traded on the internal market. (On the other hand, if the only alternative uses are to produce other goods that are traded on the internal market, then the factory manager will appropriately internalize this opportunity cost.) This could lead to suboptimal internal market outcomes.

Internal market profit: Internal market profit for a factory is the change in its internal market account balance from time $T - 1$ to time T *minus* total costs. Similarly, internal market profit for a store is the change in its account balance *plus* its total revenue. Given own period T profit $\pi_{i,T}$ and a profile of others' profits $\pi_{-i,T} = (\pi_{j,T} : j \neq i)$, participant i 's wage in period T takes

the form

$$W_{i,T}(\pi_T) = Y_i + X_i(\pi_{-i,T}) + \alpha_i \pi_{i,T} \quad (1)$$

where π_T refers to the vector of all participants' internal market profits. The Y_i term may be thought of as incorporating a base wage plus, perhaps, a term that depends on more qualitative “good citizenship” aspects of internal market performance. The $X_i(\pi_{-i,T})$ term can depend on others' profits in any way. (See the discussion in Section 4.2.) For future reference, we define

1. *Fixed wage:* $W_i = Y_i$. Employees have no incentive one way or the other to maximize their own internal market profit (or simply “own profit”).
2. *Market wage:* $W_i = Y_i + \alpha \pi_i$ for some $\alpha > 0$. Employees have a direct incentive to maximize own profit.
3. *Cooperative wage:* $W_i = Y_i + \beta \left(\pi_i + \sum_{j \neq i} \pi_j \right)$ for some $\beta > 0$. Employees have a direct incentive to maximize firm profit. (By definition, $\sum_i \pi_i$ is the same as firm profit.)
4. *Peer group relative wage:* For each factory manager f (and similarly for each store manager),

$$W_f = Y + \gamma \left(\pi_f - \frac{\sum_{f' \neq f} \pi_{f'}}{F - 1} \right)$$

Each factory manager has an incentive to increase own profit (holding the profit of his peer group fixed) and to decrease the profit of his peer group (holding own profit fixed).

3.4 No profitable trade assumption

Internal markets provide a dynamic trading environment that is very rich strategically. In real-time, each participant can place and/or remove numerous offers to buy and/or sell various quantities of all of the products at various prices. This makes the problem of characterizing

an internal market's sequential equilibria quite intractable. To make progress, our approach is to make an additional *working assumption* on internal market outcomes corresponding to the idea that there are no further profitable gains from trade.

Definition (No profitable trade property (NPT)). Trading in shares of (g, T) satisfies NPT if the final allocation of shares (at time $t = T$) is such that for all prices p and net quantity vectors $(x_{i,g,T} : i \in I)$ such that $\sum_i x_{i,g,T} = 0$, some participant i^* gets less internal market profit acquiring $x_{i^*,g,T}$ more shares at price p per share.

(A7) Trading of (g, T) shares satisfies NPT for all goods g and times T .

We shall relax assumption **(A7)** in Sections 4.1-4.2; see below.

A simple way to interpret **(A7)** is in terms of “truthful last-minute orders”. Suppose that when the last moment for trading shares of (g, T) arrives, each participant makes a set of orders corresponding to its true cost/revenue schedule for shares. For example, a factory with total cost function $TC(\cdot)$ that currently has sold (but not produced) Q shares has two kinds of truthful last-minute offers. First, for all $Q' > 0$, to buy back Q' shares for the amount that it would save by not producing these shares: $\frac{TC(Q) - TC(Q - Q')}{Q'}$ per unit for Q' units. Second, for all $Q' > 0$ to sell Q' more units for the amount that it would cost to produce them: $\frac{TC(Q + Q') - TC(Q)}{Q'}$ per unit for Q' units. If all factories (and stores) submit such last-minute orders, then **(A7)** is satisfied.

Robustness to “mistakes”. Note that **(A7)** is an assumption about the *final* allocation of shares, not about the process by which trades are made over time. For example, suppose that factory manager Frank has marginal cost \$100 and that currently shares are trading at about \$90. Although Frank is not willing to sell more shares at current prices, he decides to place a sell order at \$110, and wait and see. Unfortunately, he accidentally places the order at \$11 and

it executes immediately. If Frank could not reverse this trade, he would lose \$89 and the firm would lose \$10. (The firm would be producing a unit at cost \$100 that has opportunity cost \$90.) But since Frank’s marginal cost is \$100, he does better to re-buy the share at \$90. Frank’s loss from his mistake is still significant (\$79) but the firm’s loss disappears.

A more fundamental concern is that, in some situations, some participants may *choose* not to make profitable trades. For this reason, we relax assumption **(A7)** in our analysis. In particular, we focus on two reasons why participants may choose not to make profitable trades.¹ First, if some participant is large impact enough to have an impact on the price, he may benefit from exercising his internal market power. (See Section 4.1.) Second, if participants can coordinate their trading activities (“collude”), they may collectively benefit from refraining to make individually profitable trades. (See Section 4.2.)

4 Will internal markets maximize profits?

First we establish that the benchmark internal market maximizes firm profit given all of the assumptions of Section 3.

Theorem 1. *Given assumptions (A1-7), all internal market outcomes maximize firm profits.*

Proof. By assumption, marginal costs are increasing, marginal revenues are decreasing, there are no fixed costs, there are no externalities, and all profitable trades are made. The result then follows immediately from the First Welfare Theorem of Economics. Final internal market

¹A third reason is that, when participants have private information about their costs and revenues, they may sometimes fail to make profitable trades because of strategic posturing (Myerson and Satterthwaite (1983)). Our approach to dealing with internal market power also deals with such posturing, should it be a significant problem. See the discussion in Section 4.1.

prices constitute a price equilibrium and hence maximize total firm profits. (Each factory or store’s “welfare” is its internal market profit and the firm’s profit is the “total welfare” created by trade in the internal market. Any outcome that maximizes total welfare must therefore maximize firm profit.) □

Theorem 1 is useful as a reassuring base case but mainly as a launching pad for our analysis when its assumptions are relaxed. Not surprisingly, we will find that the benchmark internal market no longer maximizes firm profits for the same reason that a standard external market no longer maximizes total welfare. More interesting is the fact that we can propose novel solutions to the standard sorts of “market failures” that leverage the hierarchical nature of the firm. In particular, the firm has the ability to *design* the way that its employees are paid.

4.1 Internal market power and “cooperative wages”

Suppose that a factory is large relative to the overall market in that it can significantly affect internal market prices by its decision of how much quantity to produce / shares to sell. (Similar points apply to large stores.) The manager of such a factory can increase its internal market profit by committing not to make some profitable trades. For example, suppose that Frank manages a factory with zero costs and faces two store managers: Sally is willing to pay \$3 for one unit and Sven is willing to pay \$1 for one unit. If Frank can commit never to lower the price below \$3, then Sally will be willing to pay \$3. On the other hand, if Sally believes that Frank will eventually offer to sell at \$1 to make a profitable trade with Sven, then she has an incentive to wait and never pay more than \$1 herself. According to the classic “Coase conjecture”,² Frank can not credibly commit never to trade with Sven and we should expect

²As Coase (1972) noted, however, the prospect of repeated interactions across different periods may provide a credible commitment device to withhold quantity. Since internal market interactions are repeated across goods

Frank to trade with both Sally and Sven in accordance with the no profitable trade assumption **(A7)**. (See Coase (1972) and Gul, Sonnenschein, and Wilson (1986).) Nonetheless, in this section we will assume the worst: a factory manager with internal market power can and will distort production and market prices if it is in his interest to do so.

Such a factory manager will not have the incentive to distort the market if he is given a cooperative wage where his bonus depends not on his own factory's internal market profit but total firm profit (see page 10). In the extreme case, when there is just one factory and one store, both clearly have a strong common interest to determine each others' true costs and revenues and make trades exactly when they are in the best interests of the firm. Cooperative wages should be employed with care, however, since they can potentially have other sorts of unintended consequences. Two potential sorts of effects seem to us to be most noteworthy.

(1) Less powerful incentives. Say that the manager of a factory with some but not much market power earns a cooperative wage. While this manager does always have an incentive to make trades that are profitable for the firm, these incentives are fairly weak since his own trades account for only a portion of firm profit. This suggests that a hybrid sort of wage structure, including part market wages and part cooperative wages, may be better than either market wages or cooperative wages alone. Calibrating the trade-off between cooperative wages and market wages is an important topic for future experimental and field research.

(2) Potential for less orderly trading. A cooperative wage-earner (again, Frank a factory manager) has no direct interest in prices but only in firm profit. For example, suppose that Frank's marginal cost is \$150 and he knows that store manager Sally has very high marginal revenue (say \$200) but she has only made an order to buy at price \$50. Since selling a share to Sally and periods, this is a natural concern.

will increase firm profits, Frank is happy to sell a share at \$50, i.e. \$100 below cost. To put the point more broadly, a cooperative wage-earner doesn't mind being "extorted" by market wage-earners. While the final allocation may be the same as if the market wage-earners had not gamed the system in this way, the firm may have to pay more wages than it expected since market wage-earners may get very large market profits. Also, as with any gaming opportunity, employees may simply waste time and energy squeezing more money for themselves from the system.

Among the many potential approaches to address the issue of disorderly trading when some participants are cooperative wage-earners and some are not, the most promising seems to be that of endowing the cooperative wage-earners with "commitment power". In our example, suppose that Frank were able to commit only to sell at \$150, while Sally had no such commitment power. Since Frank is unable to accept any price less than \$150, Sally can not hope to get a better price.

How might cooperative wage-earners be given such power to commit while market wage-earners are not? Fortunately, as the designer of the internal market, the firm controls the *interface* through which participants interact in the internal market.

"Information interface". Imagine, for instance, that a cooperative wage-earning factory manager has a specialized "information interface" for interacting with the other participants in the internal market. This factory manager observes nothing that happens in the internal market and at each time enters only his total cost function. A robot trader automatically translates this cost information into orders to buy and/or to sell in the natural way, and the factory learns how much it must produce when trading is completed. The factory manager's bonus is still based on total firm profit but he only has an indirect role (through the costs that he announces)

in the internal market whose transactions determine that profit. Will the factory manager have the proper incentives to announce his factories' costs truthfully? “Yes” if, conditional on his announced costs, the stores will take advantage of all profitable trading opportunities, i.e. if assumption **(A7)** is satisfied in the internal market consisting only of the store managers and the robot representing the factory. In that case, announcing costs that are lower than the truth will lead to lost firm profits (and hence lost factory manager bonus) through too much production, and vice versa announcing costs that are higher than the truth will lead to lost firm profits through too little production.

Thus, it appears that the potential problems caused by participants with substantial market power in the internal market (e.g., “internal monopolists”) can be substantially avoided by giving them cooperative wages, letting them specify only their true costs, and automatically translating these costs into limit orders.

4.2 Internal market collusion and “peer group relative wages”

Coordination among participants in an internal market can be good for the firm, but collusion between them can be bad. For example, consider a simple case in which there is one factory manager and one store manager, each having private information about costs and revenues. Neither of them is sure whether they have a profitable trade and, naturally, neither of them wants to be the first to reveal its private information. The factory manager would like to “pretend” to have high costs to induce the store manager to pay a high price when he has high revenues, and vice versa the store manager would like to pretend to have low revenues. Such strategic posturing can lead to a failure to make profitable trades (Myerson and Satterthwaite (1983)). If these two worked together and found some way to credibly reveal their information and share the gains from trade, however, they and the firm would all benefit.

Indeed, the only way that store managers and factory managers working together can increase their combined internal market profit is to find new profitable opportunities to trade! For another example, consider a richer setting than our simple model in which a variety of goods produced by different factories may all be needed by a given store. That store's manager faces an "exposure problem": he prefers not to buy any of the goods until he can acquire them all at a low enough price. If the factories can coordinate and submit offers jointly, they can solve this problem for the store and then, through further trades amongst themselves, balance their share holdings with what they will produce. These examples show why a firm should hesitate before taking drastic measures to fight collusion such as prohibiting communication between internal market participants.

Harmful collusion can be defined as coordination among a "cartel" of internal market participants that (i) increases the payoffs of all members of the cartel and (ii) leads to profitable trades *not* being made. (If all profitable trades are made, no matter exactly how, Theorem 1 shows that firm profit is maximized.) As discussed above, the most pressing concern is the possibility of "peer group cartels" consisting of some factory managers or of some store managers (but not both). *If* (say) factory managers are paid a market wage, a group of them can benefit by forming a peer group cartel: they can increase their internal market profit by reducing the total quantity that they sell since this will tend to lead to an increase in the price.

What if instead they are paid peer group relative wages (see page 10)? Suppose that some group of factory managers has formed a peer group cartel. If these factory managers decide to produce the efficient quantity, nothing has changed and they get the same internal market profits and the same wage as without a cartel. If they decide to withhold some quantity, they will drive up the price and increase their internal market profits. On the other hand, the

profits of the other factories who are not in the cartel will increase even more since they are not withholding quantity, so the members of the cartel will get lower *relative* internal market profit and be worse off! (For the same reason, each individual factory manager prefers to defect from the cartel.) Similarly, providing more than the efficient quantity will make the cartel members worse off than before since then they will be providing some quantity at prices less than marginal cost.

Exclusive dealing. Given peer group relative wages, interestingly, a potential new way to collude arises between factory and store managers. For example, suppose that there are two factories f_1, f_2 and two stores s_1, s_2 . Suppose that f_1, s_1, s_2 form a cartel and they agree never to trade with factory f_2 . If they could manage this, then factory f_1 would do better than f_2 . This could potentially increase the total wages of the group $\{f_1, s_1, s_2\}$. Nonetheless, there are several difficulties associated with managing such exclusive dealing. In particular, the internal market mechanism treats all participants in the same way. So, when f_1 sells a share, it must be at a price that is less than the price at which f_2 was willing to sell a share, i.e. exactly when it is in the firm's interest for f_1 to sell a share. For this reason, we are not especially concerned about exclusive dealing.

4.3 Externalities and “taxes”

Suppose that there are externalities in sales across stores, i.e. that the amount that one store sells affects the revenues of other stores. There would be a positive externality if, for example, more sales by one store increased the visibility of the product and hence sales in other stores, or if the good enjoyed a network externality in consumption since more sales by one store would mean a larger network for customers going to other stores. There would be a negative

externality if, for example, customers have access to a liquid resale market. In that case, resale market arbitrage implies that all stores must charge the same price and more quantity sold by one store will lower the price that other stores can charge.

For the remainder of this section, we will restrict attention to the special case of negative externalities due to an active resale market. Similar ideas apply in the case of positive externalities and other sorts of negative externalities.

To make the discussion as stark as possible, suppose that the firm is a monopolist in the product market but that its individual factories and stores are so small as to view themselves as having no effect on prices (in the product market or in the internal market). In the internal market, each store has an incentive to buy shares until the internal market price equals the product market price (stores' internal market demand curve is identical to customers' demand curve in the product market). On the other hand, factories have an incentive to sell shares until the internal market price equals marginal cost. Thus, the no profitable trade assumption **(A7)** can only be satisfied when the firm produces the competitive quantity. In other words, by using an internal market the firm loses all of its monopoly power! From the firm's point of view, this is very bad. Fortunately, variations on our basic internal market design can alleviate this sort of problem.

Correcting for externalities with "taxes". Suppose that the firm imposes a "tax" τ in the internal market. For simplicity, suppose that only stores pay the tax, that all stores buy shares at the same price $p_{\tau,T}$, and that none re-sell shares.³ Given tax rate $\tau \geq 0$, each store pays an

³Our taxation scheme needs to be modified so as not to discourage trading and re-trading. One natural approach is to tax buyers on the last price that they paid for a given share. For instance, if someone buys 100 shares at \$50, then sells 50 shares at \$70, and then buys 20 shares at \$30, then he would pay tax on \$50 for the first 50 shares, pay tax on \$30 on the next 20 shares, and nothing more.

effective price $p_{\tau,T}(1 + \tau)$ while each factory gets $p_{\tau,T}$ per share. Now, store internal demand for shares crosses factories' internal supply at the quantity Q_τ such that its product market price is equal to the tax markup above marginal cost: $(1 + \tau)MC(Q_\tau) = P(Q_\tau)$. The firm's total profit then is

$$\Pi_{\tau,T} \equiv Q_\tau P(Q_\tau) - TC(Q_\tau) \quad (2)$$

The optimal tax rate is then determined implicitly by the optimal monopoly markup for the firm:

$$\tau^* = \frac{Q^* P'(Q^*)}{P(Q^*)} \quad (3)$$

where Q^* is the monopoly optimal quantity. (For instance, see Pindyck and Rubinfeld (2005), Chapter 10.) As soon as we replace the “no profitable trade” assumption **(A7)** with the corresponding “no *after-tax* profitable trade” property, the argument of Theorem 1 (after slight modification) implies that any internal market outcome with optimally chosen τ^* maximizes firm profits.

The trouble, of course, is that the firm may not be able to compute τ^* since by presumption it may not know anything about revenues until after the fact. Nonetheless, as long as the structure of revenues is not changing too rapidly, the firm can use internal market outcomes to *learn* τ^* .

Finding the optimal tax rate. The firm does not know product market inverse demand $P(\cdot)$ nor marginal costs $MC(\cdot)$ and hence can not compute the optimal tax-rate τ_T^* . Nonetheless, the firm can fine-tune its tax policy through “experimentation”, as follows. Fix $\Delta > 0$, let τ_T be the tax rate that is applicable for time T shares, and let $\Pi_{\tau_T,T}$ be the total firm profit from

production and sales at time T .

$$\text{Set } \tau_{T+1} = \tau_T + \Delta \frac{\Pi_{\tau_T, T} - \Pi_{\tau_{T-1}, T-1}}{\tau_T - \tau_{T-1}} \quad (4)$$

Suppose for the moment that product market demand and marginal costs do not change at all over time. In this case, clearly, the adjustment process (4) will lead to a tax rate that satisfies the mark-up condition (3).⁴ On the other hand, if costs and demand change very rapidly, there is little hope that the adjustment process will ever lead to an optimal tax rate. In fact, one can imagine certain situations in which demand and costs might be very negatively serially correlated in which case our adjustment process would not even tend to move the tax rate in the right direction. For instance, a Christmas tree firm will have very high sales one season and low sales in other seasons. If the firm knows that sales are seasonal, however, it could obviously apply our adjustment procedure separately to determine separate tax rates applicable in each season. And as long as year-to-year demand conditions change slowly enough, this adjustment would lead to near-optimal tax rates for each season.

4.4 Fixed costs, “conditional internal markets”, and “pivot wages”

When factories have fixed costs, the argument of Theorem 1 no longer applies since there may not be a price equilibrium. For simplicity, we will focus in this section on the case in which there is a single factory that makes a single good with a fixed cost of production FC and marginal costs $MC(Q)$. The case of multiple factories and multiple goods is discussed at the end.

Definition (Minimal average cost). Let $p_{MAC} \equiv \min_Q \frac{FC + \int_0^Q MC(x)dx}{Q}$ be the factory’s

⁴There might conceivably be more than one tax-rate satisfying this local optimality condition. As typical in optimization search algorithms, occasional large jumps in the tax rate will allow the firm to find the global optimum.

minimal average cost and $Q_{MAC} \equiv \arg \min_Q \frac{FC + \int_0^Q MC(x)dx}{Q}$ the quantity at which minimal average cost is realized (also called *minimum efficient scale*).

Since there are fixed costs, the factory willingness to supply quantity at various prices has a discontinuity at p_{MAC} . At prices $p < p_{MAC}$, there is no profitable quantity that it can sell so that the factory's "internal market supply" $S^{int}(p) = 0$. At prices $p > p_{MAC}$, on the other hand, the factory maximizes its profit by setting price equal to marginal cost, i.e. $S^{int}(p) = MC^{-1}(p)$. (In Figure 1, $S^{int}(\cdot)$ is the heavily shaded line.) There are three cases of interest, depending on the strength of internal market demand.

Case I: "Strong demand". Suppose that internal market demand is such that $D^{int}(p_{MAC}) \geq Q_{MAC}$. In this case, the only internal market outcome satisfying the no profitable trade assumption **(A7)** is for the efficient quantity $Q^* \geq Q_{MAC}$ to be produced. This maximizes firm profits. Hence, in this case the benchmark internal market leads to the optimal outcome for the firm *despite* the presence of fixed costs.

Suppose instead that $D^{int}(p_{MAC}) < Q_{MAC}$ as in Figure 1. Now, the only outcome satisfying the no profitable trade assumption is for there to be no production. In some situations, however, the firm would prefer to produce. Given production, the firm maximizes variable profit by producing Q^* in which case variable profit equals the area of triangle abc . On the other hand, the fixed cost of production equals the area of triangle bde .

Case II: "Weak demand". Suppose that $D^{int}(p_{MAC}) < Q_{MAC}$ and $AREA(abc) \leq AREA(bde)$. In this case, the internal market outcome of no production maximizes firm profits.

Case III: "Moderate demand". Suppose that $D^{int}(p_{MAC}) < Q_{MAC}$ and $AREA(abc) > AREA(bde)$. In this case, the internal market outcome of no production does not maximize firm profits. In

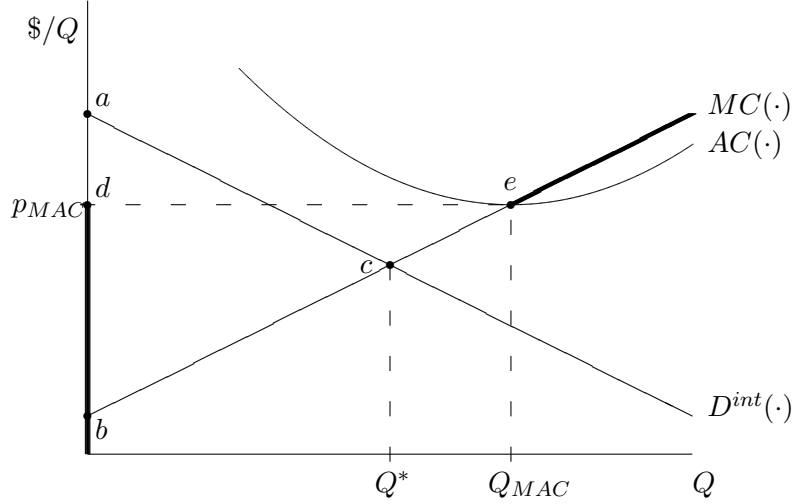


Figure 1: Given fixed costs, internal markets may lead to too little production.

particular, the firm loses $AREA(abc) - AREA(bde)$ relative to an optimal outcome.

The problem that arises given moderate demand is the same as that studied in Groves (1973) of “inducing the members of an organization to behave as if they formed a team”. When fixed costs are high enough that the factory expects negative internal market profit, the factory has the incentive to boycott production without consideration of how this impacts the stores. We will exploit Groves’ classic solution to this problem, known as the “pivot mechanism” (also called the “Groves mechanism”), in combination with a “conditional internal market”.

Conditional internal market. Trading proceeds as in a standard internal market with one extra feature. Before the production decision must be made, every participant i announces how much it would be willing to pay $\tilde{\pi}_i$ out of its internal market money account for there to be production, i.e. the factory announces $\tilde{\pi}_f$ and each store s announces $\tilde{\pi}_s$. $\tilde{\pi}_i$ may be a negative number, in which case i would be willing to pay $-\tilde{\pi}_i$ for there *not* to be production. After these numbers have been announced, the factory has discretion regarding whether to produce.⁵ If there is no

⁵An even simpler approach would be to assume that the firm can mandate whether the factory will produce.

If so, the profit-maximizing outcome will be implemented if the firm mandates production iff $\sum_i \tilde{\pi}_i > 0$. In

production, all transactions that took place in the conditional internal market are voided, i.e. all participants wind up with zero internal market profit. If there is production, on the other hand, all such transactions become binding.

Pivot wages. Unlike in the standard internal market with market wages, participant wages now depend not only on their internal market profit π_i but also on whether they were “pivotal” to the decision of whether or not to produce. In particular, define i ’s “pivot payment” p_i :

$$\begin{aligned} p_i &= \max\{0, \sum_{j \neq i} \tilde{\pi}_j\} \text{ if there is no production} \\ &= \max\{0, -\sum_{j \neq i} \tilde{\pi}_j\} \text{ if there is production} \end{aligned}$$

Definition (Pivot wage). A pivot wage for participant i is any increasing function of $\pi_i - p_i$, e.g. $W_i = \alpha(\pi_i - p_i)$.

If production is certain, a pivot wage reduces to a market wage and our conditional internal market reduces to the standard internal market. The purpose of the pivot payment is to align participants’ incentives. In the case of weak or moderate demand, for instance, the stores collectively profit from production but the factory does not. If the factory chooses not to produce, it must make a pivot payment (a “fine”) of $\sum_s \tilde{\pi}_s$. Given moderate demand but not weak demand, incurring the negative internal market profit that will result from production is worth it to avoid paying this fine.

By standard analysis of the Groves mechanism (see e.g. Mas-Colell, Whinston, and Green (1995)), each participant has a weakly dominant strategy to announce $\tilde{\pi}_i = E[\pi_i]$ where π_i is its internal market profit if production occurs (possibly random at the time of production), and

our mechanism, however, the factory also has the incentive to implement this rule so firm intervention is not necessary.

the factory has a weakly dominant strategy to produce iff $\sum_i \tilde{\pi}_i > 0$.⁶ The classic weakness of this approach is that there are potential opportunities for collusion. For example, in the case of weak demand, the stores would collectively benefit if every store s announced an outrageously high profit prediction $\tilde{\pi}_s$. (Any individual store would suffer, however, from doing so on its own.) This would not only force the factory to produce to avoid the large pivot payment but also would ensure that none of the stores would themselves be pivotal. So, the stores would get “free money” while decreasing firm profits.

Punishing inaccuracy. Fortunately, the firm has an extra lever at its disposal in our setting that is not available to Groves’ team manager. Our firm can observe ex post each participant i ’s actual internal market profit π_i . If participant i does not collude and follows its weakly dominant Groves strategy, its announcement $\tilde{\pi}_i$ should be an unbiased predictor of its actual internal market profit. To make the Groves strategy *strictly* dominant, it suffices to give each participant an incentive for accuracy. For example, consider the following form for wages:

$$W_i = \alpha(\pi_i - p_i) - \kappa(\pi_i - \tilde{\pi}_i)^2$$

Participants will then maximize expected wages by announcing $\tilde{\pi}_i = E[\pi_i]$, as needed for profit-maximization, given any $\kappa > 0$.

We have not modelled the effect of *effort* on realized costs and revenues. A natural concern might be that rewarding accuracy will create perverse incentives for employees to work more (if the prediction is too high), work less (if the prediction is too low), or work unevenly (to smooth profits across periods). All of these potential problems can be avoided, however, simply by giving every participant complete accounting flexibility when it comes to the *timing* with

⁶If participants are risk-averse, then their weakly dominant strategy is to announce a certainty equivalent, i.e. given utility $u_i(\cdot)$, $\tilde{\pi}_i$ satisfies $u_i(\tilde{\pi}_i) = E[u_i(\pi_i)]$.

which internal market profits are booked. Given that costs and revenues can be held for a while in accounting limbo, participants only need to hit their profit targets *on average* to avoid the inaccuracy punishment. This policy squarely aims the inaccuracy punishment at those whose actions disrupt the performance of the internal market, not those whose only fault is imperfect information.

Multiple factories. Suppose that there are $F > 1$ factories, each possibly with some fixed cost. Now one must determine not only whether but *where* to produce, i.e. which subset of factories to call into production. This is still a collective action problem and can be solved by the pivot mechanism. (To see how pivot payments are generalized to the case of more than two alternatives, consult Mas-Colell, Whinston, and Green (1995).) The only difference is that, to implement our approach in full generality, one must run a separate conditional internal market for every subset of factories that might be called to produce as well as elicit separate internal market profit predictions from each participant for each such subset of factories. Since there are $2^F - 1$ such subsets, this could be challenging from a practical perspective. Fortunately, in practice it will often be possible for the firm to rule out a number of possibilities a priori. For instance, suppose that the firm's factories are naturally ranked in terms of cost, i.e. factory f_i has lower total cost of producing any given quantity than factory f_j whenever $i < j$. In this case, the firm knows that the cost-minimizing production plan must involve calling factories $\{f_1, \dots, f_i\}$ for some $1 \leq i \leq F$. This means that only F conditional internal markets need to operate to maximize firm profits.

Multiple goods. Suppose in addition that these factories may produce $G > 1$ goods and that there may be economies of scope in fixed cost. Thus, for instance, one factory might only find it worthwhile to produce good B if it also produces good A . Again, this is a collective action

problem that can be solved by the pivot mechanism. The only difference now is that each conditional internal market will specify not only the set of factories that must produce but also the set of goods that each factory will produce, for $2^{F(2^G-1)} - 1$ total alternatives. “Pruning” of implausible possibilities, again, can substantially reduce the set of alternatives that must be considered.

Furthermore, there are several natural approaches to limiting the burden that being active in multiple conditional internal markets imposes on participants. Consider the “informational interface” introduced in Section 4.1. How much each store is willing to pay in the internal market for q shares is the same regardless of which factories produce. Thus, how each store would choose to bid in each conditional internal market depends only on its expected revenues and on its beliefs about future price movements in that market. While forming expectations about future price movements requires market-specific attention, trading based on whether one’s expected marginal revenue is above or below the current internal market price can be implemented automatically by a robot trader. Similarly, such a robot can generate each store’s internal market profit prediction from that store’s total revenue function. In this setup, no matter how many simultaneous conditional internal markets there may be, all the store needs to do is input its expected total revenue function. Similarly, as long as a factory’s marginal and total costs do not depend on which other factories produce, a robot trader can automate that factory’s interactions in simultaneous conditional internal markets.

5 Concluding Remarks

In most large companies today, the problem of how much capacity to allocate to different products and how to allocate and price these products for different customers is solved by a

Potential problem	Proposed solution	Where
Inexpert participants	Limited impact due to “self-correction”	Section 3.4
Internal market power	Cooperative wages	Section 4.1
Internal market collusion	Peer-group relative wages	Section 4.2
Externalities in sales	Internal market sales tax	Section 4.3
Fixed costs	Pivot wages, conditional internal market	Section 4.4

Table 1: Summary of results

time-consuming, complex, and expensive hierarchical process. In this paper, we have seen how suitably designed internal markets can solve the same problems in ways that have the potential to be more efficient, faster, and more flexible.

We have shown that under certain simple assumptions, these market solutions are provably optimal. As summarized in Table 1, we have also examined a range of potential problems that might arise in these markets and analyzed potential solutions for these problems. Interestingly, these solutions could not, in general, be used in conventional external markets because they take advantage of the fact that firms need not reward employees based on their internal market profit alone.

While we believe that much work remains to be done to actually implement internal markets like these on a large scale in real organizations, we think that the time is now ripe to begin experimenting with laboratory simulations and small-scale pilot tests of these ideas in real organizations.

References

- ACKOFF, R. (1994): *The Democratic Corporation*. Oxford University Press, New York.
- BP Climate Change Manager Jeff Morgheim (2000): “Statement to the US Senate Committee on Commerce, Science, and Transportation,” <http://www.gcric.org/OnLnDoc/pdf/pdf/morgheim000921.pdf> (accessed 6 June 2003).
- CHEN, K.-Y., L. R. FINE, AND B. A. HUBERMAN (2001): “Forecasting Uncertain Events with Small Groups,” *Proceedings of the ACM Conference on E-commerce*.
- (2003): “Predicting the Future,” *Information Systems Frontiers*, 5, 47–61.
- (2004): “Eliminating Public Knowledge Biases in Information-Aggregation Mechanisms,” *Management Science*, 50, 983–994.
- CHEN, K.-Y., AND C. R. PLOTT (1998): “Prediction Markets and Information Aggregation Mechanism: Experiments and Application,” Technical report, California Institute of Technology.
- COASE, R. (1972): “Durability and Monopoly,” *Journal of Law and Economics*, 15(1), 143–149.
- ECCLES, R. J. (1985): *The Transfer Pricing Problem: A Theory for Practice*. Lexington Books.
- GROVES, T. (1973): “Incentives in Teams,” *Econometrica*, 41(4), 617–631.
- GUL, F., H. SONNENSCHN, AND R. WILSON (1986): “Foundations of Dynamic Monopoly and the Coase Conjecture,” *Journal of Economic Theory*, 39, 155–190.
- KAMBIL, A., AND E. VAN HECK (2002): *Making Markets: How Firms Can Design and Profit from Online Auctions and Exchanges*. Harvard Business School Press, Boston, MA.

- MALONE, T. W. (2004a): “Bringing the market inside,” *Harvard Business Review*, 82(4), 106–114.
- (2004b): *The Future of Work: How the New Order of Business Will Shape Your Organization, Your Management Style, and Your Life*. Harvard Business School Press, Boston, MA.
- MAS-COLELL, A., M. WHINSTON, AND J. GREEN (1995): *Microeconomic Theory*. Oxford University Press.
- MCADAMS, D. (2005): “Storage in Internal Markets,” manuscript, available at www.mit.edu/~mcadams/papers/im/storage.pdf.
- MENDELSON, H. (1985): “Pricing computer services: queueing effects,” *Communications of the ACM*, 28(3), 312–321.
- MYERSON, R. B., AND M. A. SATTERTHWAIT (1983): “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory*, 29(2), 265–81.
- PINDYCK, R., AND D. RUBENFELD (2005): *Microeconomics*. Pearson Prentice Hall, Upper Saddle River, NJ.
- PLOTT, C. R. (2000): “Markets as Information Gathering Tools,” *Southern Economic Journal*, 67(1), 1–15.